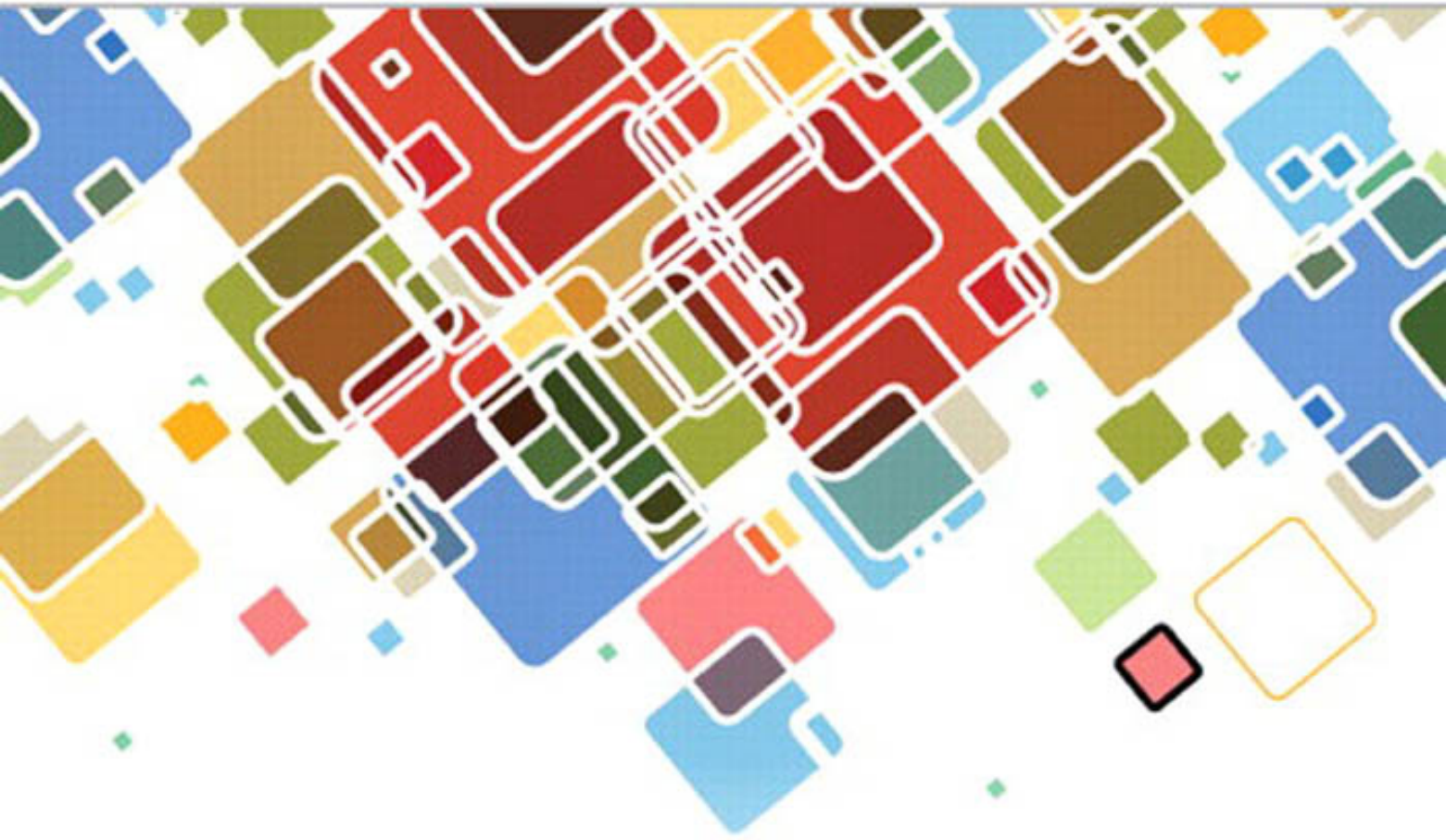
Essentials of
Database Management



Jeffrey A. Hoffer | Heikki Topi | V. Ramesh

Teradata University Network

Both PVFC databases are pre-loaded into an SQL-compliant database at the Teradata University Network. Additional production-sized sample databases are also pre-loaded. Teradata SQL Assistant allows students to run SQL commands from a Web browser from any PC. Students can create their own databases and manage usage rights for team projects. On-line SQL exams can be administered, too.

Faculty: Register at **www.teradatauniversitynetwork.com**, and then access Teradata Database under the Apply & Do menu and follow the instructions to create a course environment for your students with one or more of our textbook databases. You may also use this site to upload additional databases for your students to access. Information is available on the Teradata Database page for all the features available under Teradata SQL Assistant.

Students: Point your browser to **www.teradatauniversitynetwork.com**. The general password for student access to the Teradata University Network must be obtained from your referring professor. Access the Teradata Database under the Apply & Do link, and obtain an SQL Assistant Account for the course environment your instructor has created. Your instructor may assign a special password for this environment.

Students and faculty can conduct a trial of Teradata SQL Assistant Web Edition by clicking on Teradata SQL Assistant Trial in the Access Content category of Quick Links. This trial and associated video and documentation can work effectively with Chapter 1 of this text.

Instructors: Additional materials for faculty are available on the **Instructor Resource Center** for this book.

Instructors can register for access to faculty-only material at **www.pearsonhighered.com/irc**. Once you register, you will not have additional forms to fill out, or multiple usernames and passwords to remember in order to access new titles and/or editions. As a registered faculty member, you can login directly to download resource files.

Need help? Our dedicated Technical Support team is ready to assist instructors with questions about the media supplements that accompany this text. Visit: **http://247pearsoned.custhelp.com** for answers to frequently asked questions and toll-free user support phone numbers.

This page intentionally left blank

Essentials of Database Management

Jeffrey A. Hoffer
University of Dayton

Heikki Topi
Bentley University

V. Ramesh
Indiana University

PEARSON

Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montréal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Editor in Chief: Stephanie Wall
Executive Editor: Bob Horan
Editorial Project Manager: Kelly Loftus
Editorial Assistant: Kaylee Rotella
Director of Marketing: Maggie Moylan
Executive Marketing Manager: Anne Fahlgren
Marketing Assistant: Gianna Sandri
Senior Managing Editor: Judy Leale
Production Project Manager: Jane Bonnell
Operations Specialist: Michelle Klein
Creative Director: Jayne Conte
Cover Designer: Suzanne Duda
Cover Art: ©kras99/Fotolia
Media Project Manager, Editorial: Alana Coles
Media Project Manager, Production: Lisa Rinaldi
Composition/Full-Service Project Management: PreMediaGlobal
Printer/Binder: Edwards Brothers Malloy
Cover Printer: Lehigh-Phoenix Color/Hagerstown
Text Font: 10/12 PalatinoLTStd Roman
Video Icon: Do Ra/Fotolia

Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appear on the appropriate page within text.

Microsoft and/or its respective suppliers make no representations about the suitability of the information contained in the documents and related graphics published as part of the services for any purpose. All such documents and related graphics are provided "as is" without warranty of any kind. Microsoft and/or its respective suppliers hereby disclaim all warranties and conditions with regard to this information, including all warranties and conditions of merchantability, whether express, implied or statutory, fitness for a particular purpose, title and non-infringement. In no event shall Microsoft and/or its respective suppliers be liable for any special, indirect or consequential damages or any damages whatsoever resulting from loss of use, data or profits, whether in an action of contract, negligence or other tortious action, arising out of or in connection with the use or performance of information available from the services.

The documents and related graphics contained herein could include technical inaccuracies or typographical errors. Changes are periodically added to the information herein. Microsoft and/or its respective suppliers may make improvements and/or changes in the product(s) and/or the program(s) described herein at any time. Partial screen shots may be viewed in full within the software version specified.

Microsoft® Windows®, and Microsoft Office® are registered trademarks of the Microsoft Corporation in the U.S.A. and other countries. This book is not sponsored or endorsed by or affiliated with the Microsoft Corporation.

MySQL®, the MySQL GUI Tools® (MySQL Query Browser® and MySQL Administrator®), the MySQL Command Line Client®, the MySQL Workbench®, and the MySQL Connector/ODBC® are registered trademarks of Sun Microsystems, Inc./Oracle Corporation. Screenshots and icons reprinted with permission of Oracle Corporation. This book is not sponsored or endorsed by or affiliated with Oracle Corporation. Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Copyright © 2014 by Pearson Education, Inc. All rights reserved. Manufactured in the United States of America. This publication is protected by Copyright, and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. To obtain permission(s) to use material from this work, please submit a written request to Pearson Education, Inc., Permissions Department, One Lake Street, Upper Saddle River, New Jersey 07458, or you may fax your request to 201-236-3290.

Many of the designations by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed in initial caps or all caps.

Library of Congress Cataloging-in-Publication Data

Hoffer, Jeffrey A.
Essentials of database management / Jeffrey A. Hoffer, University of Dayton, Heikki Topi, Bentley University, V. Ramesh, Indiana University.
pages cm
ISBN-13: 978-0-13-340568-2
ISBN-10: 0-13-340568-0
1. Database management. I. Topi, Heikki. II. Ramesh, V. (Venkataraman) III. Title.
QA76.9.D3H598 2014
005.74—dc23

2013007593

10 9 8 7 6 5 4 3 2 1

PEARSON

ISBN 10: 0-13-340568-0
ISBN 13: 978-0-13-340568-2

To Dillon and Nick Domnick, Anna, Emma, and Josh Koenig—the future generation, who will have a positive impact by learning and doing.

—J.A.H.

To Anne-Louise, for her lasting love, support, and patience. To Leila and Saara, whose laughter and joy of life keep me focused on what is truly important. To my parents, Anni and Jaakko Topi, whose wise guidance and trusting encouragement are a lifelong gift.

—H.T.

To Gayathri, for her continued support and patience. To my parents, Neela and P. Venkataraman, for all their encouragement and support.

—V.R.

This page intentionally left blank

BRIEF CONTENTS

Preface xix

Part I The Context of Database Management 1

Chapter 1 The Database Environment and Development Process 2

Part II Database Analysis 43

Chapter 2 Modeling Data in the Organization 45
Chapter 3 The Enhanced E-R Model 93

Part III Database Design 119

Chapter 4 Logical Database Design and the Relational Model 121
Chapter 5 Physical Database Design and Performance 166

Part IV Implementation 189

Chapter 6 Introduction to SQL 191
Chapter 7 Advanced SQL 232
Chapter 8 Database Application Development 268
Chapter 9 Data Warehousing 300

Glossary of Acronyms 353
Glossary of Terms 355
Index 361

This page intentionally left blank

CONTENTS

Preface xix

Part I The Context of Database Management 1

An Overview of Part One 1

Chapter 1 The Database Environment and Development Process 2

Learning Objectives 2
Introduction 2
Basic Concepts and Definitions 4
- Data 4
- Data Versus Information 5
- Metadata 6

Traditional File Processing Systems 7
- Disadvantages of File Processing Systems 7
  - Program-Data Dependence 7
  - Duplication of Data 8
  - Limited Data Sharing 8
  - Lengthy Development Times 8
  - Excessive Program Maintenance 8

The Database Approach 8
- Data Models 8
  - Entities 8
  - Relationships 9
- Relational Databases 10
- Database Management Systems 10
- Advantages of the Database Approach 11
  - Program-Data Independence 11
  - Planned Data Redundancy 11
  - Improved Data Consistency 11
  - Improved Data Sharing 11
  - Increased Productivity of Application Development 12
  - Enforcement of Standards 12
  - Improved Data Quality 12
  - Improved Data Accessibility and Responsiveness 13
  - Reduced Program Maintenance 13
  - Improved Decision Support 13
- Cautions About Database Benefits 13
- Costs and Risks of the Database Approach 13
  - New, Specialized Personnel 14
  - Installation and Management Cost and Complexity 14
  - Conversion Costs 14
  - Need for Explicit Backup and Recovery 14
  - Organizational Conflict 14

Components of the Database Environment 14
The Database Development Process 16

Systems Development Life Cycle 17
PLANNING—ENTERPRISE MODELING 18
PLANNING—CONCEPTUAL DATA MODELING 18
ANALYSIS—CONCEPTUAL DATA MODELING 18
DESIGN—LOGICAL DATABASE DESIGN 18
DESIGN—PHYSICAL DATABASE DESIGN AND DEFINITION 19
IMPLEMENTATION—DATABASE IMPLEMENTATION 19
MAINTENANCE—DATABASE MAINTENANCE 19
Alternative Information Systems (IS) Development Approaches 20
Three-Schema Architecture for Database Development 21
Evolution of Database Systems 22
The Range of Database Applications 24
Personal Databases 25
Two-Tier Client/Server Databases 25
Multitier Client/Server Databases 26
Enterprise Applications 26

Developing a Database Application for Pine Valley Furniture Company 28
Database Evolution at Pine Valley Furniture Company 29
Project Planning 30
Analyzing Database Requirements 30
Designing the Database 32
Using the Database 35
Administering the Database 36
Future of Databases at Pine Valley 36
Summary 37 • Key Terms 38 • Review Questions 38 • Problems and Exercises 39 • References 40 • Further Reading 40 • Web Resources 41

Part II Database Analysis 43
An Overview of Part Two 43

Chapter 2 Modeling Data in the Organization 45
Learning Objectives 45
Introduction 45
The E-R Model: An Overview 47
Sample E-R Diagram 47
E-R Model Notation 49
Modeling the Rules of the Organization 50
Data Names and Definitions 50
DATA NAMES 51
DATA DEFINITIONS 51
GOOD DATA DEFINITIONS 51
Modeling Entities and Attributes 53
Entities 53
ENTITY TYPE VERSUS ENTITY INSTANCE 53
ENTITY TYPE VERSUS SYSTEM INPUT, OUTPUT, OR USER 54
STRONG VERSUS WEAK ENTITY TYPES 55
NAMING AND DEFINING ENTITY TYPES 55

Attributes 57
Required Versus Optional Attributes 57
Simple Versus Composite Attributes 58
Single-Valued Versus Multivalued Attributes 59
Stored Versus Derived Attributes 59
Identifier Attribute 59
Naming and Defining Attributes 61
Modeling Relationships 62
Basic Concepts and Definitions in Relationships 63
Attributes on Relationships 63
Associative Entities 65
Degree of a Relationship 66
Unary Relationship 67
Binary Relationship 67
Ternary Relationship 68
Attributes or Entity? 69
Cardinality Constraints 71
Minimum Cardinality 71
Maximum Cardinality 72
Some Examples of Relationships and Their Cardinalities 72
A Ternary Relationship 73
Modeling Time-Dependent Data 74
Modeling Multiple Relationships Between Entity Types 75
Naming and Defining Relationships 76
E-R Modeling Example: Pine Valley Furniture Company 78
Database Processing at Pine Valley Furniture 80
Showing Product Information 81
Showing Product Line Information 81
Showing Customer Order Status 82
Summary 83 • Key Terms 84 • Review Questions 84 • Problems and Exercises 85 • References 92 • Further Reading 92 • Web Resources 92

Chapter 3 The Enhanced E-R Model 93
Learning Objectives 93
Introduction 93
Representing Supertypes and Subtypes 94
Basic Concepts and Notation 94
An Example of a Supertype/Subtype Relationship 95
Attribute Inheritance 96
When to Use Supertype/Subtype Relationships 97
Representing Specialization and Generalization 97
Generalization 98
Specialization 99
Combining Specialization and Generalization 100
Specifying Constraints in Supertype/Subtype Relationships 100
Specifying Completeness Constraints 101
Total Specialization Rule 101
Partial Specialization Rule 101

Specifying Disjointness Constraints 101
DISJOINT RULE 102
OVERLAP RULE 102
Defining Subtype Discriminators 102
DISJOINT SUBTYPES 103
OVERLAPPING SUBTYPES 103
Defining Supertype/Subtype Hierarchies 105
AN EXAMPLE OF A SUPERTYPE/SUBTYPE HIERARCHY 105
SUMMARY OF SUPERTYPE/SUBTYPE HIERARCHIES 106
EER Modeling Example: Pine Valley Furniture Company 106
Packaged Data Models 110
A Revised Data Modeling Process with Packaged Data Models 110
Summary of Prepackaged Data Models 112
Summary 112 • Key Terms 113 • Review Questions 113 • Problems and Exercises 114 • References 116 • Further Reading 117 • Web Resources 117

Part III Database Design 119
An Overview of Part Three 119

Chapter 4 Logical Database Design and the Relational Model 121
Learning Objectives 121
Introduction 121
The Relational Data Model 122
Basic Definitions 122
RELATIONAL DATA STRUCTURE 123
RELATIONAL KEYS 123
PROPERTIES OF RELATIONS 124
REMOVING MULTIVALUED ATTRIBUTES FROM TABLES 124
Sample Database 124
Integrity Constraints 126
Domain Constraints 126
Entity Integrity 126
Referential Integrity 128
Creating Relational Tables 129
Well-Structured Relations 130
Transforming EER Diagrams into Relations 131
Step 1: Map Regular Entities 132
COMPOSITE ATTRIBUTES 132
MULTIVALUED ATTRIBUTES 133
Step 2: Map Weak Entities 133
WHEN TO CREATE A SURROGATE KEY 134
Step 3: Map Binary Relationships 135
MAP BINARY ONE-TO-MANY RELATIONSHIPS 135
MAP BINARY MANY-TO-MANY RELATIONSHIPS 135
MAP BINARY ONE-TO-ONE RELATIONSHIPS 136
Step 4: Map Associative Entities 137
IDENTIFIER NOT ASSIGNED 137
IDENTIFIER ASSIGNED 138

Step 5: Map Unary Relationships 139
UNARY ONE-TO-MANY RELATIONSHIPS 139
UNARY MANY-TO-MANY RELATIONSHIPS 139
Step 6: Map Ternary (and *n*-ary) Relationships 141
Step 7: Map Supertype/Subtype Relationships 141
Summary of EER-to-Relational Transformations 143
Introduction to Normalization 144
Steps in Normalization 145
Functional Dependencies and Keys 146
DETERMINANTS 146
CANDIDATE KEYS 146
Normalization Example: Pine Valley Furniture Company 148
Step 0: Represent the View in Tabular Form 148
Step 1: Convert to First Normal Form 149
REMOVE REPEATING GROUPS 149
SELECT THE PRIMARY KEY 149
ANOMALIES IN 1NF 150
Step 2: Convert to Second Normal Form 150
Step 3: Convert to Third Normal Form 151
REMOVING TRANSITIVE DEPENDENCIES 152
Determinants and Normalization 152
Merging Relations 153
An Example 154
View Integration Problems 154
SYNONYMS 154
HOMONYMS 155
TRANSITIVE DEPENDENCIES 155
SUPERTYPE/SUBTYPE RELATIONSHIPS 156
Summary 156 • Key Terms 157 • Review Questions 157 • Problems and Exercises 158 • References 165 • Further Reading 165 • Web Resources 165

Chapter 5 Physical Database Design and Performance 166
Learning Objectives 166
Introduction 166
The Physical Database Design Process 167
Physical Database Design As a Basis for Regulatory Compliance 168
Designing Fields 169
Choosing Data Types 169
CODING TECHNIQUES 170
HANDLING MISSING DATA 171
Denormalizing Data 171
Denormalization 172
OPPORTUNITIES FOR AND TYPES OF DENORMALIZATION 172
DENORMALIZE WITH CAUTION 174
Designing Physical Database Files 175
File Organizations 176
SEQUENTIAL FILE ORGANIZATIONS 178
INDEXED FILE ORGANIZATIONS 179
HASHED FILE ORGANIZATIONS 180

Designing Controls for Files 182
Using and Selecting Indexes 182
Creating a Unique Key Index 182
Creating a Secondary (Nonunique) Key Index 183
When to Use Indexes 183
Designing a Database for Optimal Query Performance 184
Summary 185 • Key Terms 185 • Review Questions 186 • Problems and Exercises 186 • References 188 • Further Reading 188 • Web Resources 188

Part IV Implementation 189
An Overview of Part Four 189

Chapter 6 Introduction to SQL 191
Learning Objectives 191
Introduction 191
Origins of the SQL Standard 192
The SQL Environment 194
Defining a Database in SQL 199
Generating SQL Database Definitions 199
Creating Tables 200
Creating Data Integrity Controls 202
Changing Table Definitions 202
Removing Tables 203
Inserting, Updating, and Deleting Data 203
Deleting Database Contents 205
Updating Database Contents 205
Internal Schema Definition in RDBMSs 205
Creating Indexes 206
Processing Single Tables 207
Clauses of the SELECT Statement 207
Using Expressions 209
Using Functions 210
Using Wildcards 213
Using Comparison Operators 213
Using Null Values 214
Using Boolean Operators 214
Using Ranges for Qualification 216
Using Distinct Values 217
Using IN and NOT IN with Lists 219
Sorting Results: The ORDER BY Clause 219
Categorizing Results: The GROUP BY Clause 220
Qualifying Results by Categories: The HAVING Clause 221
Using and Defining Views 223
Summary 225 • Key Terms 226 • Review Questions 226 • Problems and Exercises 227 • References 230 • Further Reading 230 • Web Resources 231

Chapter 7 Advanced SQL 232

Learning Objectives 232
Introduction 232
Processing Multiple Tables 233
- Equi-Join 234
- Natural Join 235
- Outer Join 236
- Sample Join Involving Four Tables 238
- Self-Join 239
- Subqueries 241
- Correlated Subqueries 246
- Using Derived Tables 247
- Combining Queries 248

Tips for Developing Queries 250
- Guidelines for Better Query Design 252

Ensuring Transaction Integrity 253
Data Dictionary Facilities 255
Triggers and Routines 256
- Triggers 256
- Routines 258
- Example Routine in Oracle's PL/SQL 259

Embedded SQL and Dynamic SQL 261

Summary 262 • Key Terms 263 • Review Questions 263 • Problems and Exercises 264 • References 267 • Further Reading 267 • Web Resources 267

Chapter 8 Database Application Development 268

Learning Objectives 268
Introduction 268
Client/Server Architectures 269
Databases in a Two-Tier Architecture 271
- A VB.NET Example 272
- A Java Example 274

Three-Tier Architectures 275
Web Application Components 278
Databases in Three-Tier Applications 278
- A JSP Web Application 279
- An ASP.NET Example 282

Key Considerations in Three-Tier Applications 283
- Stored Procedures 284
- Transactions 285
- Database Connections 285
- Key Benefits of Three-Tier Applications 285
- Cloud Computing and Three-Tier Applications 287

Extensible Markup Language (XML) 288
- Storing XML Documents 290
- Retrieving XML Documents 290

Displaying XML Data 293
XML and Web Services 293
Summary 297 • Key Terms 297 • Review Questions 298 • Problems and Exercises 298 • References 299 • Further Reading 299 • Web Resources 299

Chapter 9 Data Warehousing 300
Learning Objectives 300
Introduction 300
Basic Concepts of Data Warehousing 302
A Brief History of Data Warehousing 303
The Need for Data Warehousing 303
Need For a Company-Wide View 303
Need to Separate Operational and Informational Systems 306
Data Warehouse Architectures 306
Independent Data Mart Data Warehousing Environment 306
Dependent Data Mart and Operational Data Store Architecture: A Three-Level Approach 308
Logical Data Mart and Real-Time Data Warehouse Architecture 310
Three-Layer Data Architecture 312
Role of the Enterprise Data Model 313
Role of Metadata 313
Some Characteristics of Data Warehouse Data 314
Status Versus Event Data 314
Transient Versus Periodic Data 315
An Example of Transient and Periodic Data 315
Transient Data 315
Periodic Data 316
Other Data Warehouse Changes 317
The Derived Data Layer 317
Characteristics of Derived Data 318
The Star Schema 318
Fact Tables and Dimension Tables 319
Example Star Schema 320
Surrogate Key 321
Grain of the Fact Table 322
Duration of the Database 322
Size of the Fact Table 323
Modeling Date and Time 324
Multiple Fact Tables 324
Hierarchies 325
Slowly Changing Dimensions 328
Determining Dimensions and Facts 329
Big Data and Columnar Databases 331
If You Knew SQL Like I NoSQL 332
The User Interface 333
Role of Metadata 334
SQL OLAP Querying 334

Online Analytical Processing (OLAP) Tools 335
Slicing a Cube 336
Drill-Down 336
Summarizing More than Three Dimensions 337
Data Visualization 338
Business Performance Management and Dashboards 338
Data-Mining Tools 339
Data-Mining Techniques 340
Data-Mining Applications 340
Data Governance and Data Quality 341
Data Governance 341
Managing Data Quality 342
Characteristics of Quality Data 343
Summary 344 • Key Terms 345 • Review Questions 345 • Problems and Exercises 346 • References 350 • Further Reading 351 • Web Resources 351

Glossary of Acronyms 353
Glossary of Terms 355
Index 361

This page intentionally left blank

PREFACE

It is our great pleasure to introduce this new text in database management. This book offers its users conceptually and technically solid content that focuses on the core issues covered in most introductory data and database management courses. We pay particular attention to providing in-depth coverage of the database development life cycle that extends from conceptual data modeling using (enhanced) entity-relationship modeling through relational modeling at the logical level to database implementation and use with Structured Query Language (SQL). Within this core area, we provide in-depth coverage that is based on the latest pedagogical and technical developments in data management.

This new text, *Essentials of Database Management*, builds on the strong foundation of *Modern Database Management* (*MDM*, currently in its 11th edition) and is designed for use in introductory courses in database management that do not require the full depth and breadth of advanced topics offered by *MDM*. This goal has been achieved without sacrificing quality or rigor of coverage. Several external factors have contributed to the need for a more succinct book: Particularly in business schools, information systems programs have less real estate in the curriculum, which means that all individual topic areas, including database management, have had to let some of their content go. The percentage of non-majors in the introductory database courses has also grown, and for the non-majors the most advanced technical knowledge in areas outside the core might not be in the center of their interests. Many integrated curricula have database modules significantly shorter than a full semester; this book fits well the needs of these course units, too.

In the marketplace of concise database textbooks, this new text offers the following advantages:

- It is built on the strong foundation and extensive development process of a leading textbook (*Modern Database Management*), ensuring state-of-the-art coverage of the core database management topics.
- It strives to maintain conceptual rigor in all of its coverage.
- It provides a sharp focus on the integrated database development cycle.
- It features tools and technologies targeted for professional designers and developers (instead of focusing on the end user).
- It offers a large number of field-tested review and practice materials at various levels of difficulty.
- It is informed by many years of teaching experience at a variety of universities and involvement in leading industry and curriculum groups by the authors.

The courses that can benefit from this book are typically required as part of information systems or information technology curricula in business schools, schools of information, computer technology programs, and applied computer science departments. The Association for Information Systems (AIS), the Association for Computing Machinery (ACM), and the International Federation of Information Processing Societies (IFIPS) curriculum guidelines all outline this type of database management course. For example, the coverage of *Essentials of Database Management* is sufficient for a great majority of the core database requirements suggested in IS 2010[1], the latest information systems undergraduate model curriculum. In addition to the majors in four-year undergraduate programs, this text can serve a number of additional audiences from information and computer technology programs at community colleges to data management modules in foundation courses in graduate programs. As discussed above, it will also serve well those non-IS majors whose interests are related to the broader area of information utilization. For example, this text would form an excellent foundation for a data management course targeted to business analytics students at either undergraduate or graduate level.

[1]Topi, Heikki; Valacich, Joseph S.; Wright, Ryan T.; Kaiser, Kate; Nunamaker, Jr., Jay F.; Sipior, Janice C.; and de Vreede, Gert Jan (2010) "IS 2010: Curriculum Guidelines for Undergraduate Degree Programs in Information Systems," Communications of the Association for Information Systems: Vol. 26, Article 18.

THE KEY STRENGTHS OF *ESSENTIALS OF DATABASE MANAGEMENT*

- One of our goals is to lead other books in coverage of the latest principles, concepts, and technologies. Our experience and expertise from *Modern Database Management* gives this text a strong foundation starting with this first edition. We believe in the value of systematic, thorough, and detailed coverage of the key topics, and this book implements this principle fully. The core database development cycle content is offered with a foundational description of the database environment and a number of contextual topics that are essential for the management of organizational data, such as database development for Internet-based applications, modern data warehousing concepts, and new infrastructure technologies (such as cloud computing).
- This text focuses on what leading practitioners say is most important for database developers and savvy database users. We work with many practitioners, including the professionals of the Data Management Association (DAMA) and The Data Warehousing Institute (TDWI), leading consultants, technology leaders, and authors of articles in the most widely read professional publications. We draw on these experts to ensure that what the book includes is important and covers not only important entry-level knowledge and skills, but also those fundamentals and mindsets that lead to long-term career success.
- We present the material in a way that students find highly accessible and effective from the perspective of supporting their learning. The content, review questions, and problems and exercises in this book benefit directly from *MDM*'s continuous market feedback for 30 years. Overall, the pedagogy of the book is sound. We use many illustrations that help to make important concepts and techniques clear. We use the most modern notations. The organization of the book is flexible, so you can use chapters in whatever sequence makes sense for your students. We supplement the book with data sets to facilitate hands-on, practical learning, and with new media resources to make some of the more challenging topics more engaging.
- You may have particular interest in introducing SQL early in your course. Our text can accommodate this. First, we cover SQL in depth, devoting two full chapters to this core technology of the database field. Second, we include many SQL examples in early chapters. Third, many instructors have successfully used the two SQL chapters early in their course. Although logically appearing in the life cycle of systems development as Chapters 6 and 7, part of the implementation section of the text, many instructors have used these chapters immediately after Chapter 1 or in parallel with other early chapters. Finally, we use SQL throughout the book, for example, to illustrate Web application connections to relational databases in Chapter 8 and online analytical processing in Chapter 9.
- We have the latest in supplements and Web site support for the text. See the supplement package for details on all the resources available to you and your students.
- This text is written to be part of a modern information systems curriculum with a strong business systems development focus. Topics are included and addressed so as to reinforce principles from other typical courses, such as systems analysis and design, networking, Web site design and development, MIS principles, and computer programming. Emphasis is on the development of the database component of modern information systems and on the management of the data resource. Thus, the text is practical, supports projects and other hands-on class activities, and encourages linking database concepts to concepts being learned throughout the curriculum the student is taking. Two of the authors of this text have served in leadership roles in the development of global information systems model curricula at both undergraduate and graduate levels, and they have been able to incorporate this broad curriculum understanding in this textbook.

SPECIFIC DIFFERENCES BETWEEN THIS BOOK AND *MODERN DATABASE MANAGEMENT*

The most important difference between this book and *Modern Database Management* by the same authors is that the advanced material covered in Chapters 10–14 and Appendices A through C of *MDM* is mostly outside the scope of this text. Some key elements of Data Quality and Integration (Chapter 10 of *MDM*) and Data and Database Administration (Chapter 11 of *MDM*) are, however, introduced within Chapters 1–9 of this book. The material on Distributed Databases (Chapter 12 of *MDM*), Object-Oriented Data Modeling (Chapter 13 of *MDM*) and Using Relational Databases to Provide Object Persistence (Chapter 14 of *MDM*) is not included at all in this text. In addition, the coverage in Chapters 1–9 has been streamlined to fit better the needs of the target audience of this book, including the following changes:

- Shortening the coverage of
  - Business rules
  - Modeling time-dependent data
  - Real-time data warehousing
  - Slowly changing dimensions
  - SQL online analytical processing (OLAP) querying
- Removing the coverage of
  - Entity clustering
  - Normal forms beyond the 3rd normal form
  - Enterprise key
  - Data volume and usage analysis
  - Partitioning
  - Clustering files
  - Certain specialized SQL commands
  - Certain specialized material on views
  - Specialized coverage on certain database management systems
  - PHP
  - Factless fact tables

Our strong intent is to continue to develop both *Modern Database Management* and this text in parallel, as texts that are closely connected but that have clearly different identities and target audiences.

SUMMARY OF THE CHAPTERS

The following sections present a chapter-by-chapter description of the chapters of the book. Each chapter description presents a statement of the purpose of that chapter and a brief summary of its core content.

PART I: THE CONTEXT OF DATABASE MANAGEMENT

Chapter 1: The Database Environment and Development Process

This chapter discusses the role of databases in organizations and previews the major topics in the remainder of the text. After offering a brief introduction to the basic terminology associated with storing and retrieving data, the chapter presents a well-organized comparison of traditional file-processing systems and modern database technology. The chapter then introduces the core components of a database environment and the range of database applications that are currently in use within organizations—personal, two-tier, multi-tier, and enterprise applications. The explanation of enterprise databases includes databases that are part of enterprise resource planning systems and data warehouses. A brief history of the evolution of database technology, from pre-database files to modern object-relational technologies, is also presented. The chapter then goes on to explain the process of database development in the context of structured life cycle, prototyping, and agile methodologies. The presentation remains

consistent with the companion systems analysis texts by Hoffer, George, and Valacich. The chapter also discusses important issues in database development and frameworks for understanding database architectures and technologies (including the three-schema architecture). Reviewers frequently note the compatibility of this chapter with what students learn in systems analysis and design classes.

PART II: DATABASE ANALYSIS

Chapter 2: Modeling Data in the Organization

This chapter presents a thorough introduction to conceptual data modeling with the entity-relationship (E-R) model. The chapter title emphasizes the reason for the entity-relationship model: to unambiguously document the rules of the business that influence database design. Specific subsections explain in detail how to name and define elements of a data model, which are essential in developing an unambiguous E-R diagram. The chapter proceeds from simple to more complex examples, and it concludes with a comprehensive E-R diagram for the Pine Valley Furniture Company.

Chapter 3: The Enhanced E-R Model

This chapter presents a discussion of several advanced E-R data model constructs. The chapter particularly focuses on providing a thorough coverage of supertype/subtype relationships and includes a comprehensive example of an extended E-R data model for the Pine Valley Furniture Company.

PART III: DATABASE DESIGN

Chapter 4: Logical Database Design and the Relational Model

This chapter describes the process of converting a conceptual data model to a relational data model, as well as discusses how to merge new relations into an existing normalized database. It provides a conceptually sound and practically relevant introduction to normalization, emphasizing the importance of the use of functional dependencies and determinants as the basis for normalization. The chapter also features a discussion of the characteristics of foreign keys. The chapter emphasizes the basic concepts of the relational data model and the role of the database designer in the logical design process.

Chapter 5: Physical Database Design and Performance

This chapter describes the steps that are essential in achieving an efficient database design, with a strong focus on those aspects of database design and implementation that are typically within the control of a database professional in a modern database environment. The chapter has an emphasis on ways to improve database performance, with references to specific techniques available in Oracle and other database management systems (DBMSs) to improve database processing performance. The discussion of indexes includes descriptions of the types of indexes that are widely available in database technologies as techniques to improve query processing speed.

PART IV: IMPLEMENTATION

Chapter 6: Introduction to SQL

This chapter presents a thorough introduction to the SQL used by most DBMSs (SQL:1999) and introduces the changes that are included in the latest proposed standard (SQL:2008). The coverage of SQL is extensive and divided between this and the next chapter. This chapter includes examples of SQL code, using mostly SQL:1999 and SQL:2008 syntax, as well as some Oracle 11g and Microsoft SQL Server syntax. Some unique features of MySQL are mentioned. Dynamic views are also covered. Chapter 6 explains the SQL commands needed to create and maintain a database and to program single-table queries. Coverage of dual-table, IS NULL/IS NOT NULL, more built-in functions, derived tables, and rules for aggregate functions and the GROUP BY clause

is included. The chapter also uses the Pine Valley Furniture Company case to illustrate a wide variety of practical queries and query results.

Chapter 7: Advanced SQL

This chapter continues the description of SQL, with a careful explanation of multiple-table queries, transaction integrity, data dictionaries, triggers and stored procedures (including a clear explanation of the differences between these two), and embedded SQL in other programming language programs. The OUTER JOIN command is also covered. Standard SQL is also used in Chapter 7. This chapter illustrates how to store the results of a query in a derived table and the CAST command to convert data between different data types. A section on self-joins is included, as is an explanation of when to use EXISTS (NOT EXISTS) versus IN (NOT IN). The chapter also covers subqueries and correlated subqueries, two of the most complex and powerful constructs in SQL.

Chapter 8: Database Application Development

This chapter provides a modern discussion of the concepts of client/server architecture and applications, middleware, and database access in contemporary database environments. Technologies that are commonly used to create two- and three-tier applications are presented. The chapter includes sample application programs that demonstrate how to access databases from popular programming languages such as Java, VB.NET, ASP.NET, and JSP. The chapter also presents expanded coverage of the role of Extensible Markup Language (XML) and related technologies in data storage and retrieval. Topics covered include basics of XML schemas, XQuery, and XSLT. The chapter concludes with an overview of Web services, associated standards and technologies, and their role in seamless, secure movement of data in Web-based applications. A brief introduction to service-oriented architecture (SOA) is also presented.

Chapter 9: Data Warehousing

This chapter describes the basic concepts of data warehousing, the reasons data warehousing is regarded as critical to competitive advantage in many organizations, and the database design activities and structures unique to data warehousing. The chapter also reviews best practices for determining requirements for a dimensional model and introduces the emerging column databases technology, which has been developed especially for data warehousing applications. The chapter includes material on important new approaches to and technologies supporting data management that are currently generating a high level of interest, including the big data concept and associated tools and techniques. Exercises provide hands-on practice with a data mart, using SQL and a BI tool called MicroStrategy that is available on Teradata University Network. Additional topics include alternative data warehouse architectures and the dimensional data model (or star schema) for data warehouses. A deep explanation of how to handle slowly changing dimensional data is provided. Operational data store; independent, dependent, and logical data marts; and various forms of online analytical processing (OLAP) are defined. User interfaces, including OLAP, data visualization, business performance management and dashboards, and data mining are also described. Brief coverage of data quality and governance is also included in this chapter.

PEDAGOGY

This text includes a broad range of teaching resources to support a variety of pedagogical approaches. The chapters are sufficiently independent so that they can be used in a variety of sequences, depending on the instructor's preference. As discussed earlier, some instructors might prefer to introduce SQL and hands-on database work earlier in the course than the chapter sequence suggests; others might decide to wait to cover physical database design and database implementation at the end of the course.

The chapters include several standard features that support the teaching and learning processes in a consistent way:

1. Learning objectives appear at the beginning of each chapter, as a preview of the major concepts and skills students will learn from that chapter. The learning objectives also provide a great study review aid for students as they prepare for assignments and examinations.
2. Chapter introductions and summaries both encapsulate the main concepts of each chapter and link material to related chapters, providing students with a comprehensive conceptual framework for the course.
3. Each chapter includes a review section, which features Key Terms, Review Questions, and Problems and Exercises. These elements will be discussed in more detail below.
4. A running glossary defines key terms in the page margins as they are discussed in the text. These terms are also defined at the end of the text, in the Glossary of Terms. Also included is the end-of-book Glossary of Acronyms for abbreviations commonly used in database management.

The end-of-chapter materials include the following elements:

A. ***Key Terms*** A key terms list provides an easy reference to the students with which they can test their grasp of important concepts, basic facts, and significant issues.
B. ***Review Questions*** Each chapter includes a rich variety of review questions that cover the key content of the chapter and provide the students with a comprehensive tool for ensuring that they have mastered the conceptual content of the chapter.
C. ***Problems and Exercises*** All chapters include a comprehensive set of Problems and Exercises that are presented in the increasing order of difficulty, making it easier for instructors and students to find the appropriate material. The Problems and Exercises focus on systematic development of the skills featured in each chapter. In several chapters, the data sets provided for the text are used extensively.
D. ***Web Resources*** Each chapter contains a list of updated and validated URLs for Web sites that contain information that supplements the chapter. These Web sites cover online publication archives, vendors, electronic publications, industry standards organizations, and many other sources. These sites allow students and instructors to find updated product information, innovations that have appeared since the printing of the book, background information to explore topics in greater depth, and resources for writing research papers.

We encourage instructors to customize their use of this book to meet the needs of both their curriculum and student career paths. The modular nature of the text, its broad coverage, extensive illustrations, and its inclusion of advanced topics and emerging issues make customization easy. The many references to current publications and Web sites can help instructors develop supplemental reading lists or expand classroom discussion beyond material presented in the text.

SUPPLEMENTS: WWW.PEARSONHIGHERED.COM/HOFFER

A comprehensive and flexible technology support package is available to enhance the teaching and learning experience. All instructor and student supplements are available on the text Web site: ***www.pearsonhighered.com/hoffer***.

For Students

The following online resources are available to students:

- The *Web Resources* module includes the Web links referenced at the end of each chapter in the text to help students further explore database management topics on the Web.
- A full *glossary* is available, along with a glossary of acronyms.

- *Links to sites where students can use our data sets* are provided. Although our data sets are provided in formats that are easily loaded on computers at your university or on student PCs, some instructors will not want the responsibility of supporting local data sets. The application service providers with whom we have developed arrangements (e.g., ***www.teradatauniversitynetwork.com***) provide thin-client interfaces to SQL coding environments. See the text's Web site and the inside front cover for more details.
- *Accompanying databases* are also provided. Two versions of the Pine Valley Furniture Company case have been created and populated. One version is scoped to match the textbook examples. A second version is fleshed out with more data and tables, as well as sample forms, reports, and modules coded in Visual Basic. This version is not complete, however, so that students can create missing tables and additional forms, reports, and modules. Databases are provided in several formats (ASCII tables, Oracle script, and Microsoft Access), but formats vary for the two versions. Some documentation of the databases is also provided. Both versions of the PVFC database are also provided on Teradata University Network.
- *Several custom-developed short videos that address key concepts and skills from different sections of the book* help students learn material that may be more difficult to understand. The videos integrate exercises and mini lectures.

For Instructors

The following online resources are available to instructors:

- The *Instructor's Resource Manual* provides chapter-by-chapter instructor objectives, classroom ideas, and answers to Review Questions and Problems and Exercises.
- The *Test Bank* and *TestGen* include a comprehensive set of test questions in multiple-choice, true/false, and short-answer format, ranked according to level of difficulty and referenced with page numbers and topic headings from the text. The Test Bank is available in Microsoft Word and as the Computerized TestGen. TestGen is PC/Mac compatible and preloaded with all of the test bank questions. You can manually or randomly view test questions and drag and drop to create a test. You can also add or modify test bank questions as needed.
- *PowerPoint Presentation slides* feature lecture notes that highlight key terms and concepts. Instructors can customize the presentation by adding their own slides or editing existing ones.
- The *Image Library* is a collection of the text art organized by chapter. It includes all figures, tables, and screenshots and can be used to enhance class lectures and PowerPoint slides.

ACKNOWLEDGEMENTS

Every book project is a collaborative effort to which numerous people contribute in addition to the authors. We are deeply grateful to all those who supported and enabled the preparation of this first edition of *Essentials of Database Management*. We particularly appreciate Executive Editor Bob Horan's guidance and his willingness to work with us to develop the concept for this book. The resulting identity is different from both *MDM* and competing textbooks, and we believe the book will serve a broad range of users well. Equally important are the contributions of the faculty members around the world who have used *MDM* and provided us with highly valuable feedback that has enabled the continuous improvement of *MDM* and motivated the development of this book. We would like to give special thanks to those colleagues who supported the process of developing the core ideas underlying this text and gave direct and insightful feedback, including Hillol Bala, Indiana University; Sue Brown, University of Arizona; Monica Garfield, Bentley University; Vijay Khatri, Indiana University; Ryan Wright, University of Massachusetts, Amherst; and Jennifer Xu, Bentley University.

We also wish to thank our reviewers for their detailed suggestions and insights: Jason Chen, Gonzaga University; Jinzhu Gao, University of Minnesota, Morris;

Rick Gibson, American University; William Hochstettler, Franklin University; Barbara Klein, University of Michigan-Dearborn; Alan Labouseur, Marist College; Lara Preiser-Huoy, California State Polytechnic University, Pomona; and Uzma Raja, University of Alabama.

We are also grateful to the staff and associates of Pearson Education for their support and guidance throughout this project. In addition to Executive Editor Bob Horan, we would like to thank Senior Editorial Project Manager Kelly Loftus, who kept us on track and made sure everything was complete; Production Project Manager Jane Bonnell; Executive Marketing Manager Anne Fahlgren; and Editorial Assistant Kaylee Rotella. Our special thanks go to Katy Gabel at PreMediaGlobal, whose supervision of the production process was excellent.

Finally, we give immeasurable thanks to our spouses, who were willing to not only accept but also support the additional requirements of this new project. In particular, we continue to marvel at the commitment of Patty Hoffer, who has lived the lonely life of a textbook author's spouse through 11 editions of *MDM* over more than 30 years in addition to this new text. We also want to offer our sincere thanks to Anne-Louise Klaus for her strong support for Heikki's involvement in this project. Similarly, Gayathri Mani's continued support and understanding are very much appreciated. Much of the value of this text is due to their patience, encouragement, and love; we alone bear the responsibility for any errors or omissions.

Jeffrey A. Hoffer

Heikki Topi

V. Ramesh

PART I

The Context of Database Management

AN OVERVIEW OF PART ONE

In this chapter and opening part of the book, we set the context and provide basic database concepts and definitions used throughout the text. In this part, we portray database management as an exciting, challenging, and growing field that provides numerous career opportunities for information systems students. Databases continue to become a more common part of everyday living and a more central component of business operations. From the database that stores contact information in your personal digital assistant (PDA) or smartphone to the very large databases that support enterprise-wide information systems, databases have become the central points of data storage that were envisioned decades ago. Customer relationship management and Internet shopping are examples of two database-dependent activities that have developed in recent years. The development of data warehouses that provide managers the opportunity for deeper and broader historical analysis of data also continues to take on more importance.

We begin by providing basic definitions of *data*, *database*, *metadata*, *database management system*, *data warehouse*, and other terms associated with this environment. We compare databases with the older file management systems they replaced and describe several important advantages that are enabled by the carefully planned use of databases.

The chapter in this introductory part of the book also describes the general steps followed in the analysis, design, implementation, and administration of databases. Further, this chapter also illustrates how the database development process fits into the overall information systems development process. Database development for both structured life cycle and prototyping methodologies is explained. We introduce enterprise data modeling, which sets the range and general contents of organizational databases. This is often the first step in database development. We introduce the concept of schemas and the three-schema architecture, which is the dominant approach in modern database systems. We describe the major components of the database environment and the types of applications, as well as two-tier, multitier, and enterprise databases. Enterprise databases include those that are used to support enterprise resource planning systems and data warehouses. Finally, we describe the roles of the various people who are typically involved in a database development project. The Pine Valley Furniture Company case is introduced and used to illustrate many of the principles and concepts of database management. This case is used throughout the text as a continuing example of the use of database management systems.

Chapter 1

The Database Environment and Development Process

CHAPTER 1

The Database Environment and Development Process

Visit www.pearsonhighered.com/hoffer to view the accompanying video for this chapter.

LEARNING OBJECTIVES

After studying this chapter, you should be able to:

- Concisely define each of the following key terms: **database, data, information, metadata, database application, data model, entity, relational database, database management system (DBMS), data independence, user view, database administration, data administration, constraint, computer-aided software engineering (CASE) tools, repository, enterprise data modeling, systems development life cycle (SDLC), conceptual schema, logical schema, physical schema, prototyping, agile software development, enterprise resource planning (ERP),** and **data warehouse.**
- Name several limitations of conventional file processing systems.
- Explain at least 10 advantages of the database approach, compared to traditional file processing.
- Identify several costs and risks of the database approach.
- List and briefly describe nine components of a typical database environment.
- Identify four categories of applications that use databases and their key characteristics.
- Describe the life cycle of a systems development project, with an emphasis on the purpose of database analysis, design, and implementation activities.
- Explain the prototyping and agile-development approaches to database and application development.
- Explain the roles of individuals who design, implement, use, and administer databases.
- Explain the differences among external, conceptual, and internal schemas and the reasons for the three-schema architecture for databases.

INTRODUCTION

Over the past two decades, there has been enormous growth in the number and importance of database applications. Databases are used to store, manipulate, and retrieve data in nearly every type of organization, including business, health care, education, government, and libraries. Database technology is routinely used by individuals on personal computers, by workgroups accessing databases on network servers, and by employees using enterprise-wide distributed applications.

Databases are also accessed by customers and other remote users through diverse technologies, such as automated teller machines, Web browsers, smartphones, and intelligent living and office environments. Most Web-based applications depend on a database foundation.

Following this period of rapid growth, will the demand for databases and database technology level off? Very likely not! In the highly competitive environment of today, there is every indication that database technology will assume even greater importance. Managers seek to use knowledge derived from databases for competitive advantage. For example, detailed sales databases can be mined to determine customer buying patterns as a basis for advertising and marketing campaigns. Organizations embed procedures called *alerts* in databases to warn of unusual conditions, such as impending stock shortages or opportunities to sell additional products, and to trigger appropriate actions.

Although the future of databases is assured, much work remains to be done. Many organizations have a proliferation of incompatible databases that were developed to meet immediate needs rather than based on a planned strategy or a well-managed evolution. Enormous amounts of data are trapped in older, "legacy" systems, and the data are often of poor quality. New skills are required to design and manage data warehouses and to integrate databases with Internet applications. There is a shortage of skills in areas such as database analysis, database design, data administration, and database administration. We address these and other important issues in this textbook to equip you for the jobs of the future.

A course in database management has emerged as one of the most important courses in the information systems curriculum today. Many schools have added an additional elective course in data warehousing or database administration to provide in-depth coverage of these important topics. As information systems professionals, you must be prepared to analyze database requirements and design and implement databases within the context of information systems development. You also must be prepared to consult with end users and show them how they can use databases (or data warehouses) to build decision support systems and executive information systems for competitive advantage. And the widespread use of databases attached to Web sites that return dynamic information to users of these sites requires that you understand not only how to link databases to the Web-based applications but also how to secure those databases so that their contents can be viewed, but not compromised, by outside users.

In this chapter, we introduce the basic concepts of databases and database management systems (DBMSs). We describe traditional file management systems and some of their shortcomings that led to the database approach. Next, we consider the benefits, costs, and risks of using the database approach. We review the range of technologies used to build, use, and manage databases; describe the types of applications that use databases—personal, two-tier, three-tier, and enterprise; and describe how databases have evolved over the past five decades.

Because a database is one part of an information system, this chapter also examines how the database development process fits into the overall information systems development process. The chapter emphasizes the need to coordinate database development with all the other activities in the development of a complete information system. It includes highlights from a hypothetical database development process at Pine Valley Furniture Company. Using this example, the chapter introduces tools for developing databases on personal computers and the process of extracting data from enterprise databases for use in stand-alone applications.

There are several reasons for discussing database development at this point. First, although you may have used the basic capabilities of a database management system, such as Microsoft Access, you may not yet have developed an understanding of how these databases were developed. Using simple examples, this chapter briefly illustrates what you will be able to do after you complete a database course using this text. Thus, this chapter helps you develop a vision and context for each topic developed in detail in subsequent chapters.

Second, many students learn best from a text full of concrete examples. Although all of the chapters in this text contain numerous examples, illustrations, and actual database designs and code, each chapter concentrates on a specific aspect of database management. We have designed this chapter to help you understand, with minimal technical details, how all of these individual aspects of database management are related and how database development tasks and skills relate to what you are learning in other information systems courses.

Finally, many instructors want you to begin the initial steps of a database development group or individual project early in your database course. This chapter gives you an idea of how to structure a database development project sufficient to begin a course exercise. Obviously, because this is only the first chapter, many of the examples and notations we will use will be much simpler than those required for your project, for other course assignments, or in a real organization.

One note of caution: You will not learn how to design or develop databases just from this chapter. Sorry! We have purposely kept the content of this chapter introductory and simplified. Many of the notations used in this chapter are not exactly like the ones you will learn in subsequent chapters. Our purpose in this chapter is to give you a general understanding of the key steps and types of skills, not to teach you specific techniques. You will, however, learn fundamental concepts and definitions and develop an intuition and motivation for the skills and knowledge presented in later chapters.

BASIC CONCEPTS AND DEFINITIONS

Database
An organized collection of logically related data.

We define a **database** as an organized collection of logically related data. Not many words in the definition, but have you looked at the size of this book? There is a lot to do to fulfill this definition.

A database may be of any size and complexity. For example, a salesperson may maintain a small database of customer contacts—consisting of a few megabytes of data—on her laptop computer. A large corporation may build a large database consisting of several terabytes of data (a *terabyte* is a trillion bytes) on a large mainframe computer that is used for decision support applications (Winter, 1997). Very large data warehouses contain more than a petabyte of data. (A *petabyte* is a quadrillion bytes.) (We assume throughout the text that all databases are computer based.)

Data

Historically, the term *data* referred to facts concerning objects and events that could be recorded and stored on computer media. For example, in a salesperson's database, the data would include facts such as customer name, address, and telephone number. This type of data is called *structured* data. The most important structured data types are numeric, character, and dates. Structured data are stored in tabular form (in tables, relations, arrays, spreadsheets, etc.) and are most commonly found in traditional databases and data warehouses.

The traditional definition of data now needs to be expanded to reflect a new reality: Databases today are used to store objects such as documents, e-mails, maps, photographic images, sound, and video segments in addition to structured data. For example, the salesperson's database might include a photo image of the customer contact. It might also include a sound recording or video clip about the most recent product. This type of data is referred to as *unstructured* data, or as multimedia data. Today structured and unstructured data are often combined in the same database to create a true multimedia environment. For example, an automobile repair shop can combine structured data (describing customers and automobiles) with multimedia data (photo images of the damaged autos and scanned images of insurance claim forms).

Data
Stored representations of objects and events that have meaning and importance in the user's environment.

An expanded definition of **data** that includes structured and unstructured types is "a stored representation of objects and events that have meaning and importance in the user's environment."

Data Versus Information

The terms *data* and *information* are closely related and in fact are often used interchangeably. However, it is useful to distinguish between data and information. We define **information** as data that have been processed in such a way that the knowledge of the person who uses the data is increased. For example, consider the following list of facts:

Information
Data that have been processed in such a way as to increase the knowledge of the person who uses the data.

Baker, Kenneth D.	324917628
Doyle, Joan E.	476193248
Finkle, Clive R.	548429344
Lewis, John C.	551742186
McFerran, Debra R.	409723145

These facts satisfy our definition of data, but most people would agree that the data are useless in their present form. Even if we guess that this is a list of people's names paired with their Social Security numbers, the data remain useless because we have no idea what the entries mean. Notice what happens when we place the same data in a context, as shown in Figure 1-1a.

By adding a few additional data items and providing some structure, we recognize a class roster for a particular course. This is useful information to some users, such as the course instructor and the registrar's office. Of course, as general awareness of the importance of strong data security has increased, few organizations still use Social Security numbers as identifiers. Instead, most organizations use an internally generated number for identification purposes.

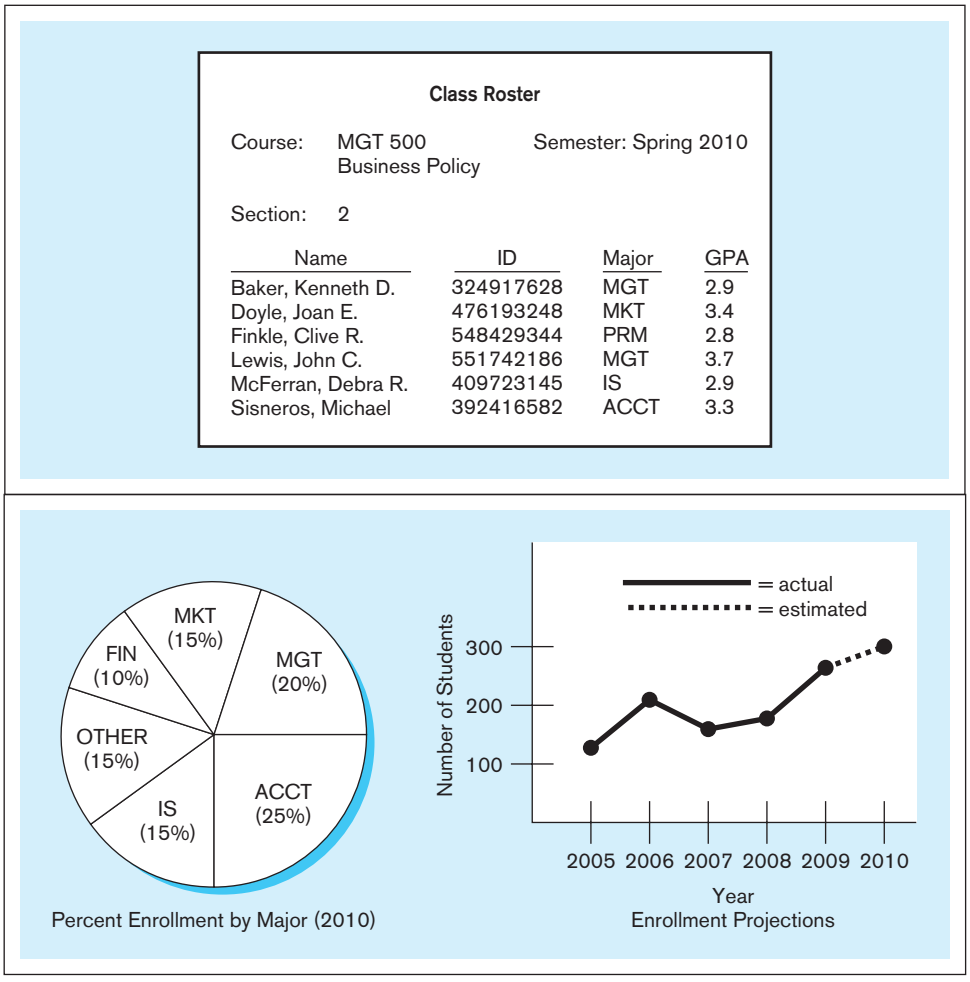


FIGURE 1-1 Converting data to information
(a) Data in context

(b) Summarized data

Another way to convert data into information is to summarize them or otherwise process and present them for human interpretation. For example, Figure 1-1b shows summarized student enrollment data presented as graphical information. This information could be used as a basis for deciding whether to add new courses or to hire new faculty members.

In practice, according to our definitions, databases today may contain either data or information (or both). For example, a database may contain an image of the class roster document shown in Figure 1-1a. Also, data are often preprocessed and stored in summarized form in databases that are used for decision support. Throughout this text, we use the term *database* without distinguishing its contents as data or information.

Metadata

Metadata

Data that describe the properties or characteristics of end-user data and the context of those data.

As we have indicated, data become useful only when placed in some context. The primary mechanism for providing context for data is metadata. **Metadata** are data that describe the properties or characteristics of end-user data and the context of that data. Some of the properties that are typically described include data names, definitions, length (or size), and allowable values. Metadata describing data context include the source of the data, where the data are stored, ownership (or stewardship), and usage. Although it may seem circular, many people think of metadata as "data about data."

Some sample metadata for the Class Roster (Figure 1-1a) are listed in Table 1-1. For each data item that appears in the Class Roster, the metadata show the data item name, the data type, length, minimum and maximum allowable values (where appropriate), a brief description of each data item, and the source of the data (sometimes called the *system of record*). Notice the distinction between data and metadata. Metadata are once removed from data. That is, metadata describe the properties of data but are separate from that data. Thus, the metadata shown in Table 1-1 do not include any sample data from the Class Roster of Figure 1-1a. Metadata enable database designers and users to understand what data exist, what the data mean, and how to distinguish between data items that at first glance look similar. Managing metadata is at least as crucial as managing the associated data because data without clear meaning can be confusing, misinterpreted, or erroneous. Typically, much of the metadata are stored as part of the database and may be retrieved using the same approaches that are used to retrieve data or information.

Data can be stored in files or in databases. In the following sections, we examine the progression from file processing systems to databases and the advantages and disadvantages of each.

TABLE 1-1 Example Metadata for Class Roster

Data Item			Metadata			
Name	**Type**	**Length**	**Min**	**Max**	**Description**	**Source**
Course	Alphanumeric	30			Course ID and name	Academic Unit
Section	Integer	1	1	9	Section number	Registrar
Semester	Alphanumeric	10			Semester and year	Registrar
Name	Alphanumeric	30			Student name	Student IS
ID	Integer	9			Student ID (SSN)	Student IS
Major	Alphanumeric	4			Student major	Student IS
GPA	Decimal	3	0.0	4.0	Student grade point average	Academic Unit

TRADITIONAL FILE PROCESSING SYSTEMS

PINE VALLEY FURNITURE

When computer-based data processing was first available, there were no databases. To be useful for business applications, computers had to store, manipulate, and retrieve large files of data. Computer file processing systems were developed for this purpose. Figure 1-2 shows an example of three computer applications based on the file processing approach used at Pine Valley Furniture Company. The systems shown are Order Filling, Invoicing, and Payroll. The figure also shows the major data files associated with each application.

A file is a collection of related records. Figure 1-2 shows, for example, that the Order Filling System has three files: Customer Master, Inventory Master, and Back Order. Notice that there is duplication of some of the files used by the three applications, which is typical of file processing systems.

As business applications became more complex, it became evident that traditional file processing systems had a number of shortcomings and limitations (described next). As a result, these systems have been replaced by database processing systems in most business applications today. Nevertheless, you should have at least some familiarity with the disadvantages of file processing systems since understanding the problems and limitations inherent in file processing systems can help you avoid the same problems when designing database systems.

Disadvantages of File Processing Systems

Several disadvantages associated with conventional file processing systems are listed in Table 1-2 and described briefly next.

TABLE 1-2 Disadvantages of File Processing Systems

Program-data dependence
Duplication of data
Limited data sharing
Lengthy development times
Excessive program maintenance

PROGRAM-DATA DEPENDENCE File descriptions are stored within each **database application** program that accesses a given file. For example, in the Invoicing System in Figure 1-2, Program A accesses the Inventory Pricing File and the Customer Master File. Because the program contains a detailed file description for these files, any change to a file structure requires changes to the file descriptions for all programs that access the file.

Database application

An application program (or set of related programs) that is used to perform a series of database activities (create, read, update, and delete) on behalf of database users.

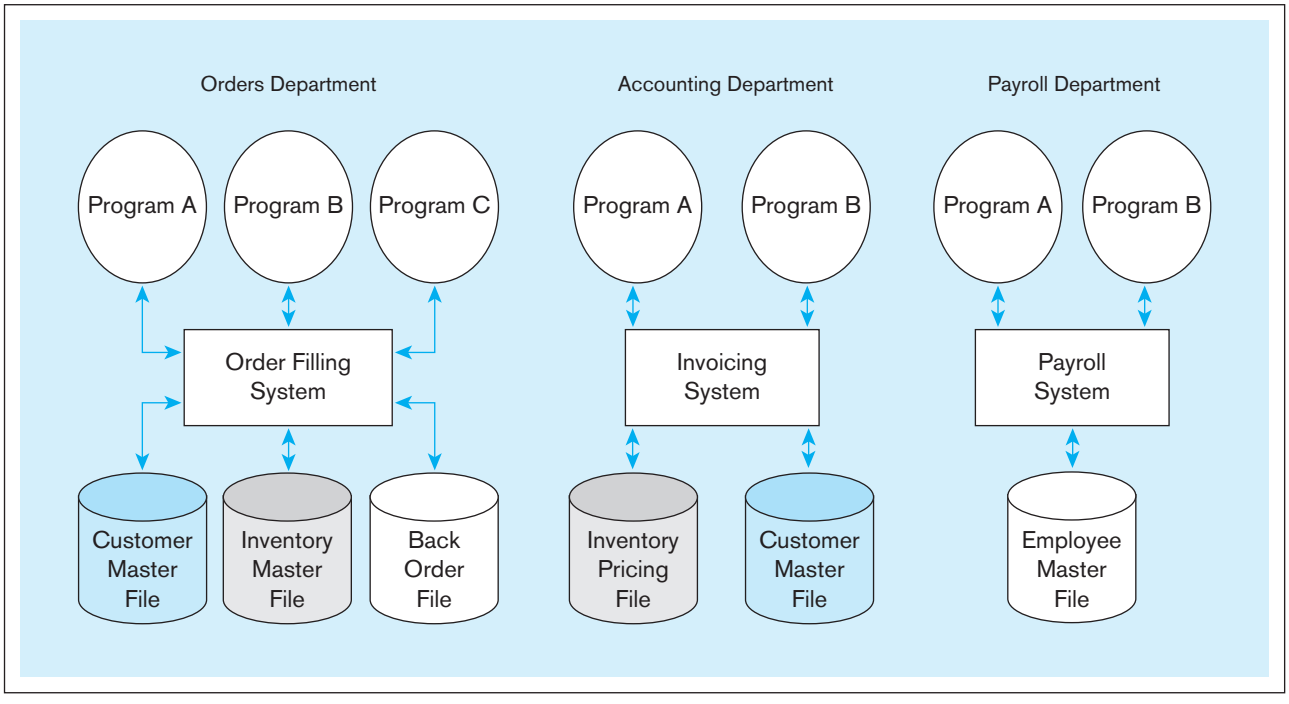


FIGURE 1-2 Old file processing systems at Pine Valley Furniture Company

DUPLICATION OF DATA Because applications are often developed independently in file processing systems, unplanned duplicate data files are the rule rather than the exception. For example, in Figure 1-2, the Order Filling System contains an Inventory Master File, whereas the Invoicing System contains an Inventory Pricing File. These files contain data describing Pine Valley Furniture Company's products, such as product description, unit price, and quantity on hand. This duplication is wasteful because it requires additional storage space and increased effort to keep all files up to date.

LIMITED DATA SHARING With the traditional file processing approach, each application has its own private files, and users have little opportunity to share data outside their own applications. Notice in Figure 1-2, for example, that users in the Accounting Department have access to the Invoicing System and its files, but they probably do not have access to the Order Filling System or to the Payroll System and their files.

LENGTHY DEVELOPMENT TIMES With traditional file processing systems, each new application requires that the developer essentially start from scratch by designing new file formats and descriptions and then writing the file access logic for each new program. The lengthy development times required are inconsistent with today's fast-paced business environment, in which time to market (or time to production for an information system) is a key business success factor.

EXCESSIVE PROGRAM MAINTENANCE The preceding factors all combined to create a heavy program maintenance load in organizations that relied on traditional file processing systems. In fact, as much as 80 percent of the total information system's development budget might be devoted to program maintenance in such organizations. This in turn means that resources (time, people, and money) are not being spent on developing new applications.

One way to overcome the limitations identified above is by using the database approach, which, as explained in the next section, is as much a way to manage organizational data as it is a set of technologies for defining, creating, maintaining, and using these data.

THE DATABASE APPROACH

So, how do we overcome the flaws of file processing? No, we don't call Ghostbusters, but we do something better: We follow the database approach. We first begin by defining some core concepts that are fundamental in understanding the database approach to managing data. We then describe how the database approach can overcome the limitations of the file processing approach.

Data Models

Data model
Graphical systems used to capture the nature and relationships among data.

Designing a database properly is fundamental to establishing a database that meets the needs of the users. **Data models** capture the nature of and relationships among data and are used at different levels of abstraction as a database is conceptualized and designed. The effectiveness and efficiency of a database is directly associated with the structure of the database. Various graphical systems exist that convey this structure and are used to produce data models that can be understood by end users, systems analysts, and database designers. Chapters 2 and 3 are devoted to developing your understanding of data modeling. A typical data model is made up entities, attributes, and relationships and the most common data modeling representation is the entity-relationship model. A brief description is presented next. More details will be forthcoming in Chapters 2 and 3.

Entity
A person, a place, an object, an event, or a concept in the user environment about which the organization wishes to maintain data.

ENTITIES Customers and orders are objects about which a business maintains information. They are referred to as "entities." An **entity** is like a noun in that it describes a person, a place, an object, an event, or a concept in the business environment for which

information must be recorded and retained. CUSTOMER and ORDER are entities in Figure 1-3a. The data you are interested in capturing about the entity (e.g., Customer Name) is called an *attribute*. Data are recorded for many customers. Each customer's information is referred to as an *instance* of CUSTOMER.

RELATIONSHIPS A well-structured database establishes the *relationships* between entities that exist in organizational data so that desired information can be retrieved. Most relationships are one-to-many (*1*:*M*) or many-to-many (*M*:*N*). A customer can place (the Places relationship) more than one order with a company. However, each order is usually associated with (the Is Placed By relationship) a particular customer. Figure 1-3a shows the *1*:*M* relationship of customers who may place one or more orders; the *1*:*M* nature of the relationship is marked by the crow's foot attached to the rectangle (entity) labeled ORDER. This relationship appears to be the same in Figures 1-3a and 1-3b. However, the relationship between orders and products is *M*:*N*. An order may be

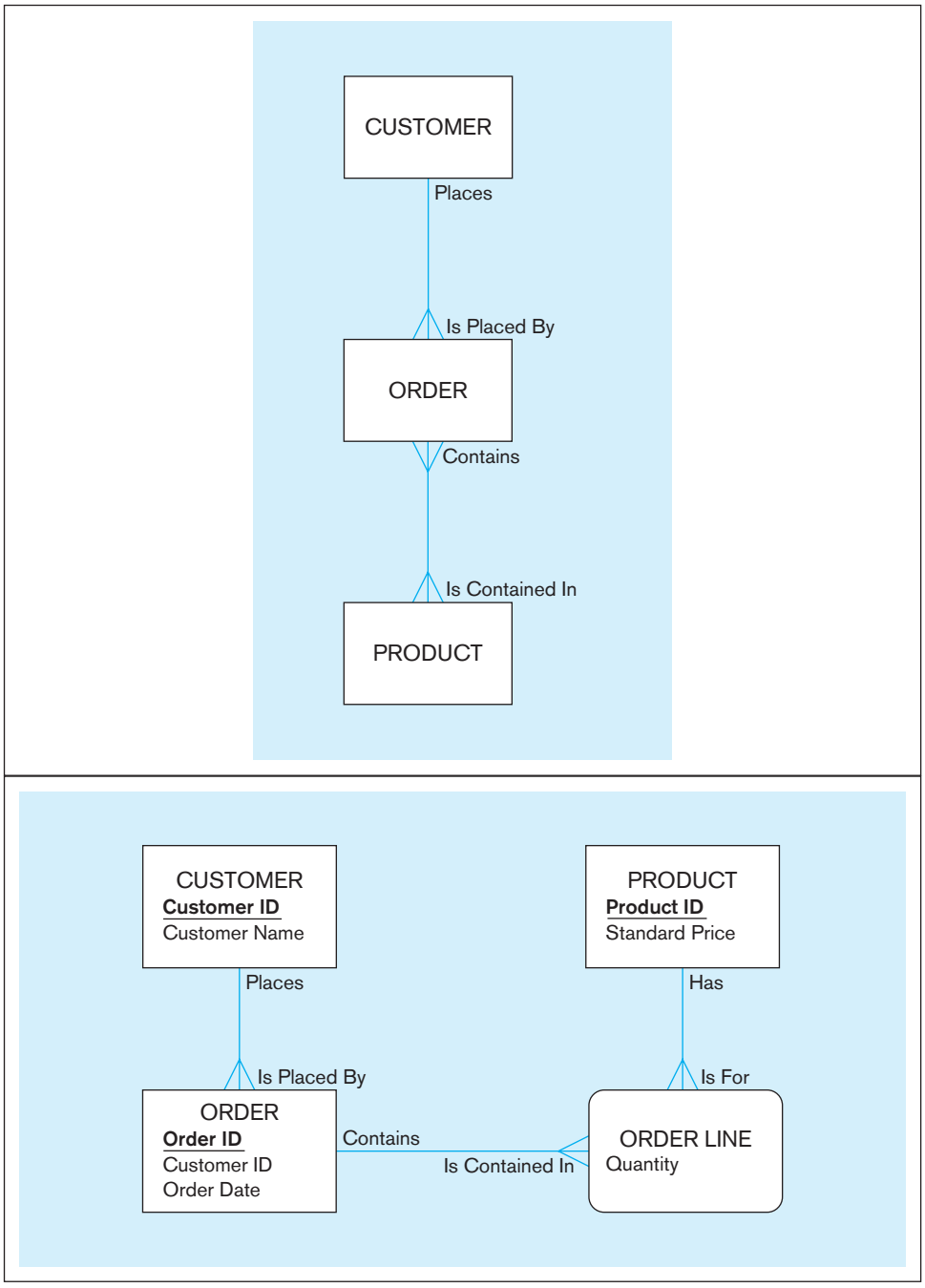


FIGURE 1-3 Comparison of enterprise and project-level data models
(a) Segment of an enterprise data model

(b) Segment of a project data model